Ultrasonographic analysis in knee osteoarthritis: evaluation of inter-observer reliability

K. Bevers[1], M.C. Zweers[1], C.H.M. van den Ende[1], H.A. Martens[1], E. Mahler[1], J.W.J. Bijlsma[2], R.J. Wakefield[3], F.H. van den Hoogen[1], A.A. den Broeder[1]

[1]Department of Rheumatology, Sint Maartenskliniek, Nijmegen, The Netherlands; [2]Department of Rheumatology and Clinical Immunology, University Medical Centre Utrecht, Utrecht, The Netherlands; [3]Section of Musculoskeletal Disease, Leeds Institute of Molecular Medicine, University of Leeds, Leeds, United Kingdom.

Abstract

Objective

Evidence for the validity of US in detecting structural joint pathology in OA is increasing. However, despite the rapidly emerging field of US in OA, few studies have reported on the inter-observer reliability of US to date. The objective of this study was to assess inter-observer reliability of ultrasonography (US) in the evaluation of specifically defined features in osteoarthritis (OA) of the knee.

Methods

US was performed independently by two rheumatologists in 60 outpatients fulfilling the American College of Rheumatology clinical criteria for knee OA. The acquisition protocol comprised medial meniscus protrusion, synovial hypertrophy, effusion, infrapatellar bursitis and cartilage thickness. Cartilage thickness and meniscal protrusion (if >3 mm) were measured on a continuous scale, all other variables were scored dichotomously.

Results

Inter-observer agreement (κ-value) was moderate for protrusion of the medial meniscus (0.54), good for infrapatellar bursitis (0.66) and effusion (0.74), excellent for Bakers' cyst (0.85) and poor for the detection of synovial hypertrophy (-0.08). Inter-observer reliability was good for the measurement of medial meniscus protrusion (correlation coefficient 0.80, 95% limits of agreement -1.93 to 1.94 mm) and cartilage thickness (correlation coefficient 0.62 and 0.68, 95% limits of agreement -0.87 to 0.84 mm and -0.77 to 0.96 mm at the medial and lateral condyle respectively).

Conclusion

This study demonstrated good reproducibility of US in the assessment of the majority of the investigated mechanical, inflammatory and degenerative features of knee OA, and contributes to exploring the use of US in knee OA as a useful tool in research as well as in clinical practice.

Key words

osteoarthritis, ultrasonography, inter-observer variation, knee

Karen Bevers, MD
Manon C. Zweers, MD, PhD
Cornelia H.M. van den Ende, PhD
Henk A. Martens, MD, PhD
Elien Mahler, MD
Johannes W.J. Bijlsma, MD, PhD
Richard J. Wakefield, MD, PhD
Frank H. van den Hoogen, MD, PhD
Alfons A. den Broeder, MD, PhD

Please address correspondence to:
Dr Karen Bevers,
Department of Rheumatology,
Sint Maartenskliniek Nijmegen,
P.O. Box 9011,
6500 GM Nijmegen, The Netherlands.
E-mail: k.bevers@maartenskliniek.nl

Received on June 19, 2011; accepted in revised form on December 11, 2011.

Competing interests: none declared.

INTRODUCTION

Osteoarthritis (OA) is a common joint disorder, with the knee being one of the most frequently involved sites. Knee OA has traditionally been imaged with conventional radiography, providing little information about soft tissue structures (1). As OA is a disease of the entire joint, characterised by cartilage breakdown, subchondral bone alterations, formation of osteophytes, meniscal degeneration and synovial inflammation, information about these pathologic findings will likely give insight into the complex process of development and progression of knee OA (2). Ultrasonography (US), a non-invasive, safe and relatively inexpensive imaging tool, allows visualisation of these structures (3, 4). For US to be implemented structurally, guidelines on the use of US in OA are being developed (5-9).

Evidence for the validity of US in detecting structural pathology in OA is increasing (10). Cartilage degeneration, a hallmark of OA, can be accurately imaged with US as good correlations between US and histology have been found (10-12). Furthermore, good agreement between US and magnetic resonance imaging (MRI) in visualising effusion and synovial hypertrophy in patients with symptomatic knee OA has been shown (13).

However, despite the rapidly emerging field of US in OA, few studies have reported on the inter-observer reliability of US to date. As US is known to be an operator-dependant modality, lack of inter-reader agreement could restrict its use, especially in clinical practice. So far, few studies have reported reproducibility data on US in knee OA and no standardised and reproducible US protocol for knee OA has been developed. Although various studies report US detection of synovial inflammation (14), there is a lack of reliability data on synovial hypertrophy and effusion in knee OA. In studies on patients with knee arthritis inter-observer agreement for synovial hypertrophy varied from 0.4 to 0.7 and reported agreement on the presence or absence of effusion in the knee varied between 0.65 and 0.77 (12). Data on inter-observer agreement for cartilage thickness included only a few patients and inter-rater reliability ranged from 0.6 to 0.9 (15, 16) Furthermore, only a limited number of studies addressed interobserver reliability of US in the detection of meniscus lesions, Baker's cyst and infrapatellar bursa (17, 18). Therefore, the aim of this study was to investigate the interobserver variability of a set of specifically defined US features comprising inflammatory, degenerative and mechanical aspects in knee OA.

Patients and methods

Patients

A total of 60 consecutive patients attending our outpatient clinic and fulfilling the American College of Rheumatology (ACR) (19) clinical criteria for knee OA were included in our study. The criteria used were: knee pain (>15 days of the last month) plus at least three of the following: age >50 years, morning stiffness <30 minutes, crepitus, bony enlargement, bony tenderness, no palpable warmth. Exclusion criteria were: inflammatory rheumatic diseases or deposition diseases possibly leading to secondary OA, severe co-morbidity exceeding the complaints of knee OA and planned orthopaedic surgery within the next 3 months. Patients underwent weight bearing antero-posterior radiographs of the knee, and had ultrasound assessment of the most symptomatic knee by two trained ultrasonographers. We studied the most symptomatic knee because this study was carried out in the framework of an osteoarthritis cohort which focuses on an index knee. The local Medical Research Ethics Committee, region Arnhem-Nijmegen (The Netherlands) approved the study design (study number 2009/095).

US investigation

All patients were assessed independently on the same day by two rheumatologists trained and certified in US (HM and KB). They had 4 and 2 years respectively professional experience in US and had performed >500 US investigations each. To guarantee independency of observations, the US investigators were blinded to the results of prior US-, x-ray- or physical examinations.

The two rheumatologists involved reached consensus on the US acquisition prior to the study by investigating 15 patients together. Based on previous US studies, EULAR guidelines(5) and pathophysiologic concepts of OA, we decided to investigate 6 different US features:

1. Effusion: a ≥4mm anechoic area in the suprapatellar recess, evaluated using a longitudinal scan in line with the patellar tendon with the leg in passive full extension. Structures are labelled as effusion if they are fully compressible to discriminate effusion from synovial hypertrophy. The site of maximal effusion is measured.
2. Synovial hypertrophy: a hypoechoic area (which is poorly compressible and nondisplaceable) of ≥2mm in the suprapatellar recess in line with the patellar tendon, measured with the leg in full extension with a longitudinal scan. The site of maximal hypertrophy is measured.
3. Meniscal protrusion: protrusion of meniscal tissue out of the joint space >3 mm from the joint line, evaluated at the medial joint space with the knee in full extension with a longitudinal scan (Fig. 1A). The maximal protrusion is measured from the joint line to the menical-synovial fluid interface. It is measured with the medial collateral ligament in sight, perpendicular to the joint line.
4. Deep infrapatellar bursitis: an enlarged infrapatellar bursa (>2 mm) on both longitudinal and transverse scans with the knee in 45° flexion.
5. Baker's cyst: a hypo-anechoic area between the medial gastrocnemius and the semimembranosus tendon examined with the patient in prone position on the posterior/medial side of the fully extended knee applying a transverse and longitudinal scan. The maximum diameter is measured (mm) in a transverse plane.
6. Femoral cartilage thickness: an anechoic band with sharp hyperechoic margins, measured perpendicular to the surface at the intercondylar notch and at the medial and lateral condyle (5 mm just medial or lateral from the top of the condyle), with the transducer immediately above the patella in a transverse plane and with the knee in maximum flexion (Fig. 1B). The outer hyperechoic margin is included in the measurement

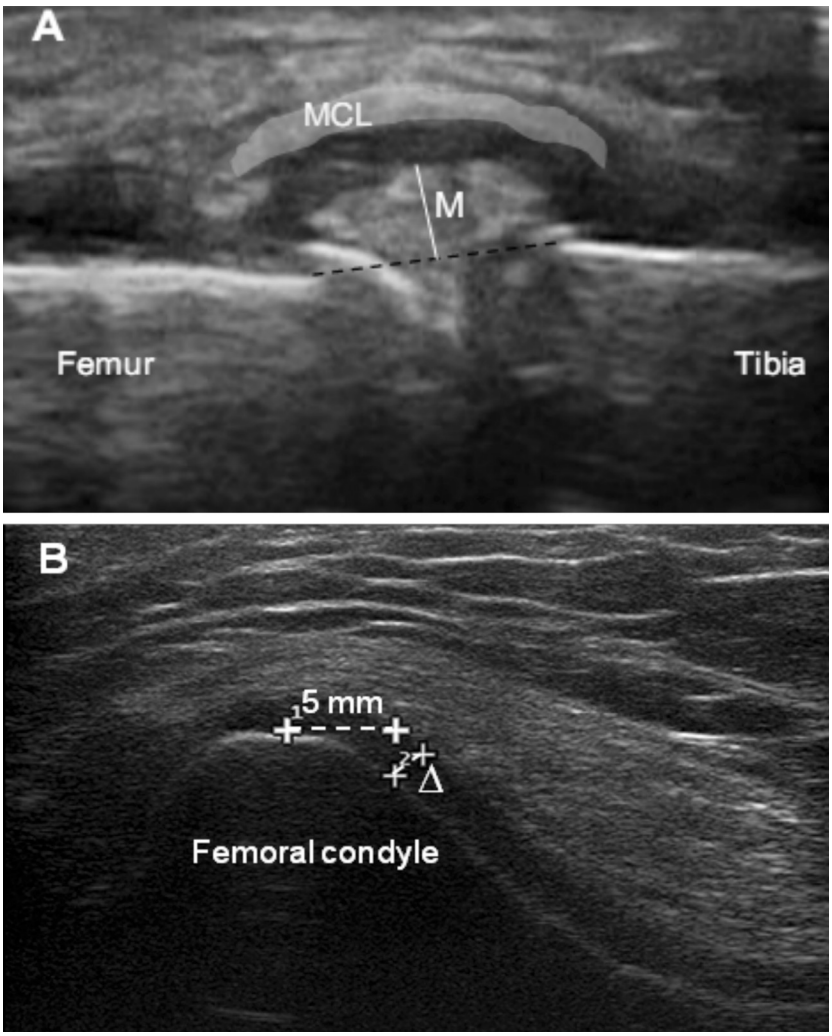

Fig. 1. A. Ultrasound image of medial meniscus with measurement of protrusion. Protrusion of the meniscus (M) was measured between the medial collateral ligament (MCL) and the joint space (dashed line).
B. Femoral condyle cartilage with measurement of cartilage thickness of lateral condyle. Cartilage thickness was measured perpendicular to the surface 5 mm from the top of the condyle (Δ).

The ultrasound machine used in this study was a MyLab 25 gold (Esaote Biomedica, Genoa, Italy), with a 35 mm linear transducer (frequency 8-15 mHz). The complete US investigation took about five minutes per patient.

Statistical analysis

Based on kappa statistics, sixty patients were required in this study to reach an agreement of 0.7 (95%-confidence interval (CI) of 0.2), assuming a prevalence of 15–50% of different US features. Inter-observer agreement for dichotomous variables was evaluated using unweighted kappa statistics and percentage of exact agreement. Kappa values <0.20 were considered poor, between 0.20 and 0.40 fair, between 0.41 and 0.60 moderate, between 0.61 and 0.80 good and >0.80 excellent (20).

Inter-observer agreement for continuous or ordinal variables was assessed by calculating the concordance correlation coefficient and respective 95% CI (21). Bland Altman analysis was performed to determine 95% levels of agreement and modified Bland Altman plots using regression analyses to account for trend were depicted (22). Statistical analysis was performed using the statistical software package Stata10 (StataCorp, Texas, USA).

Table I. Patient characteristics (n=60).

Women, %	72
Age (years), mean (SD)	53.6 (10.3)
BMI, mean (SD)	29.1 (6.3)
Disease duration (years), median (range)	3 (0.1–22)
Duration of complaints (years), median (range)	8 (0.2–40)
K&L score (%)	
1	28
2	43
3	17
4	12
Most symptomatic knee (n)	
Right	29
Left	31

Table II. Prevalence and inter-observer reliability of ultrasonography-detected pathology in knee osteoarthritis.

Observation	Prevalence KB (%)	Prevalence HM (%)	Agreement (%)	Kappa (95% CI)
Meniscus protrusion	68	68	80	0.54 (0.31–0.77)
Baker's cyst	22	29	95	0.85 (0.68–1.00)
Infrapatellar bursitis	2	3	98	0.66 (0.04–1.00)
Effusion	8	15	95	0.74 (0.47–1.00)
Synovial hypertrophy	10	5	83	-0.08 (-0.16–0.01)

Table III. Inter-observer reliability for measurement of femoral articular cartilage thickness and meniscal protrusion.

Observation	Correlation coefficient (95% CI)	Difference; limits of agreement (95% CI)	
Medial femoral cartilage thickness	0.62 (0.46–0.79)	-0.11	-0.98–0.77
Lateral femoral cartilage thickness	0.68 (0.54–0.82)	0.30	-1.31–1.92
Intercondylar notch cartilage thickness	0.50 (0.33–0.66)	-0.05	-0.85–0.75
Medial meniscus protrusion	0.80 (0.68–0.92)	0.19	-3.48–3.78

Results

Sixty patients were included in our study. Patient characteristics are shown in Table I. The prevalence of the various pathophysiological features differed notably; infrapatellar bursitis and synovial hypertrophy were observed infrequently, whereas meniscal protrusion was found in over half of the patients (Table II). Inter-observer agreement was poor for the presence or absence of synovial thickening, moderate for protrusion of the medial meniscus and good for infrapatellar bursitis and joint effusion and excellent for Bakers' cyst (Table II).

Cartilage thickness (mean±SD) was 1.93 (0.54) mm (range 0.0–3.8) at the medial condyle, 1.99 (0.59) mm (range 0.0–3.6) at the lateral condyle and 2.47 (0.68) mm (range 1.6–5.0) at the intercondylar notch. Inter-observer agreements for femoral articular cartilage thickness measurements were moderate to good (Table III). Figure 2A-C shows the difference between the 2 observers' measurements of cartilage thickness and mean measurements. The 95% limits of agreement as determined by Bland Altman analysis were as follows: medial condyle -0.87 to 0.84 mm; lateral condyle -0.77 to 0.96 mm; intercondylar notch -1.53 to 0.99 mm.

When protrusion of the medial meniscus was present, inter-observer agreement for the degree of bulging was excellent (Table III, 95% limits of agreement -1.93 to 1.94 mm). However, with increasing size of meniscal protrusion, the measurement became less precise (Fig. 2D).

Discussion

Although US is a highly operator-dependant technique, few studies have addressed the inter-observer variability of US in the assessment of knee pathology in OA. Knowledge of US reliability is pivotal before US can be implicated in research or clinical practice. In this study, we assessed the inter-observer reliability of multiple concurrently studied US features in the evaluation of OA. We chose six different US features, covering inflammatory, mechanical and degenerative aspects of knee OA. Our results show moderate to excellent inter-observer reliability for detection of infrapatellar bursitis, effusion, meniscal protrusion and Baker's cysts and for the measurement of cartilage thickness.

Despite standardised measurements and the use of bony/anatomic landmarks, we found lower inter-observer reliability for cartilage thickness than previously demonstrated for healthy subjects (11, 16, 23, 24). This is most likely due to the difficulty of defining the boundaries of cartilage-bone and cartilage-soft tissue at sites of cartilage damage (12, 16, 25). In addition, maximal flexion of the knee (as opposed to fixed flexion), required to visualise the weight-bearing parts of the femoral condyles, potentially differs between measurements due to pain during knee bending. Furthermore, the standardisation of our measurement resulted in suboptimal isonisation of the ultrasound beam, which can result in underestimation of the cartilage thickness.

In contrast to studies of inflammatory arthropathies (14), inter-observer reliability was poor for the detection of synovial hypertrophy in osteoarthritic knees. This is probably due to the low occurrence of synovial hypertrophy in our population and hence limited US training and consensus on acquisition between the observers. In a large cohort of painful knee osteoarthritis, prevalence of synovial hypertrophy and effusion on US was 16.9% and 43.7%, respectively (7). In this study, inflammation correlated strongly with advanced radiographic disease. Therefore, this higher incidence of inflammatory characteristics might reflect more advanced disease, (67% of patients with K&L score of ≥3) (7) as compared to our cohort (29% of patients with K&L score ≥3). In addition to this, we might

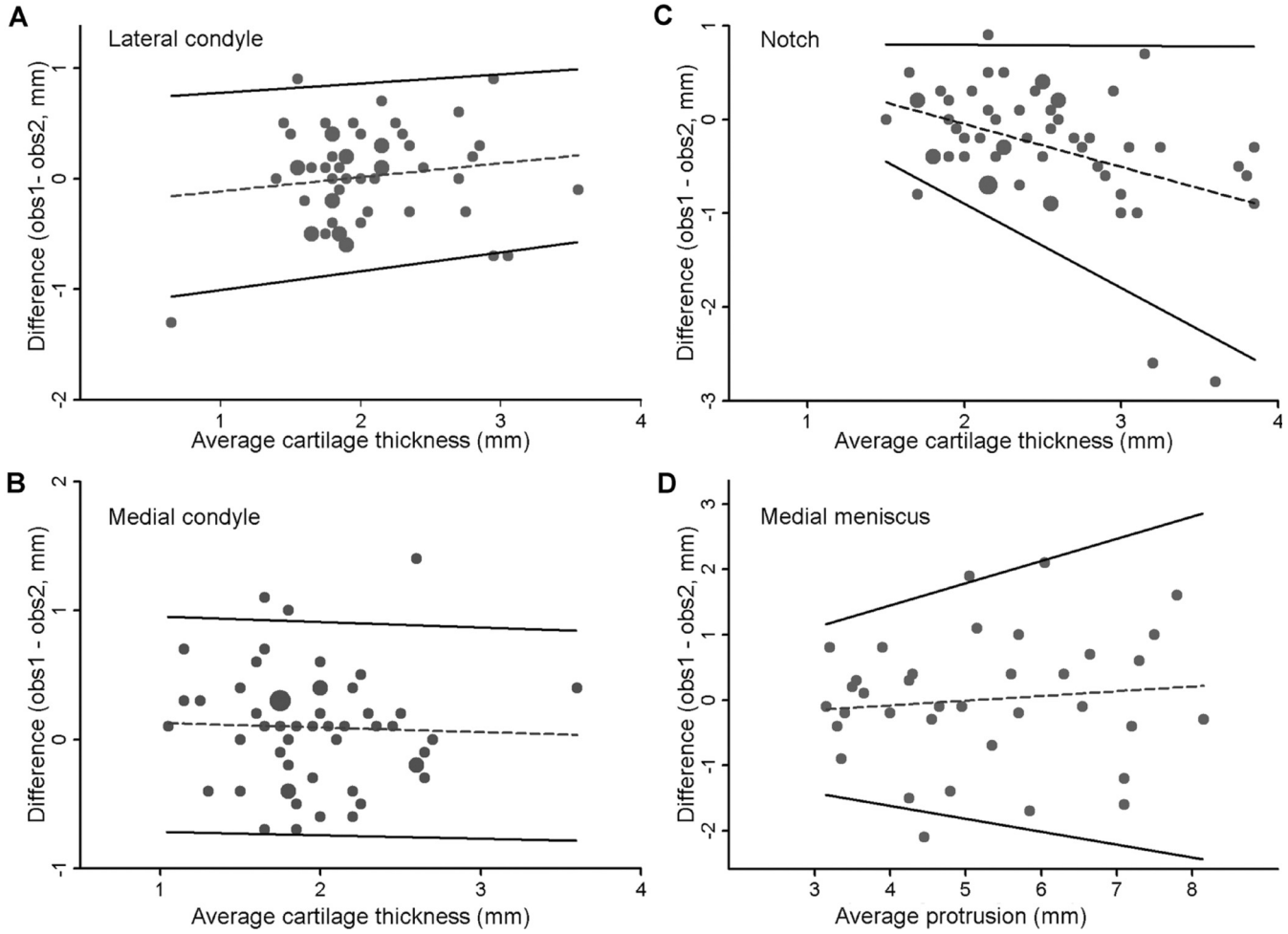

Fig. 2. Adapted Bland-Altman plots illustrating inter-observer agreement for US measurement of femoral cartilage thickness (**A**-**C**) and medial meniscus protrusion (**D**). Dashed line: mean, solid line: 95% limit of agreement.

have missed a number of patients with effusion, as we measured fluid only at the suprapatellar recess with the leg relaxed. Tension on the quadriceps muscles might reveal more fluid from the suprapatellar pouch.

The main limitation of our study was the relatively small study population, owing to which different US features were observed in a small number of patients and precision was sometimes lower than aimed for. Furthermore, we did not study intra-observer reproducibility because of practical difficulties (particularly blinding for previous measurements) and because intra-observer reliability is generally higher than inter-observer reliability. Some previous studies have determined inter-observer reliability of reading acquired ultrasound images, however, as ultrasound is a dynamic investigation, it is more important to study differences in the acquisition of images. This study contributes to exploring the use of US in knee OA, by addressing the visualisation of a set of soft tissue structures in the knee. In conclusion, the present study demonstrates moderate to excellent inter-observer reliability of US in the inspection of different inflammatory, mechanical and degenerative characteristics of knee OA. Therefore, US potentially might prove to become a useful tool in research as well as in clinical practice.

References

1. BIJLSMA JW, KNAHR K: Strategies for the prevention and management of osteoarthritis of the hip and knee. *Best Pract Res Clin Rheumatol* 2007; 21: 59-76.
2. HUNTER DJ: Imaging insights on the epidemiology and pathophysiology of osteoarthritis. *Rheum Dis Clin North Am* 2009; 35: 447-63.
3. MEENAGH G, FILIPPUCCI E, DELLE SEDIE A *et al*.: Ultrasound imaging for the rheumatologist XXX. Sonographic assessment of the painful knee. *Clin Exp Rheumatol* 2010; 28: 803-5.
4. SCIRE CA, IAGNOCCO A, MEENAGH G *et al*.: Ultrasound imaging for the rheumatologist. XXVIII. Impact of sonographic knee joint involvement in recent-onset inflammatory polyarthritis. *Clin Exp Rheumatol* 2010; 28: 449-53.
5. BACKHAUS M, BURMESTER GR, GERBER T *et al*.: Guidelines for musculoskeletal ultrasound in rheumatology. *Ann Rheum Dis* 2001; 60: 641-9.
6. CONAGHAN P, D'AGOSTINO MA, RAVAUD P *et al*.: EULAR report on the use of ultrasonography in painful knee osteoarthritis. Part 2: exploring decision rules for clinical utility. *Ann Rheum Dis* 2005; 64: 1710-4.
7. D'AGOSTINO MA, CONAGHAN P, LE BM *et al*.: EULAR report on the use of ultrasonography in painful knee osteoarthritis. Part 1: prevalence of inflammation in osteoarthritis. *Ann Rheum Dis* 2005; 64: 1703-9.
8. D'AGOSTINO MA, CONAGHAN PG, NAREDO E *et al*.: The OMERACT ultrasound task force -- Advances and priorities. *J Rheumatol* 2009; 36: 1829-32.
9. WAKEFIELD RJ, BALINT PV, SZKUDLAREK M *et al*.: Musculoskeletal ultrasound includ-

ing definitions for ultrasonographic pathology. *J Rheumatol* 2005; 32: 2485-7.
10. KEEN HI, WAKEFIELD RJ, CONAGHAN PG: A systematic review of ultrasonography in osteoarthritis. *Ann Rheum Dis* 2009; 68: 611-9.
11. MARTINO F, ETTORRE GC, ANGELELLI G *et al*.: Validity of echographic evaluation of cartilage in gonarthrosis. Preliminary report. *Clin Rheumatol* 1993; 12: 178-83.
12. MCCUNE WJ, DEDRICK DK, AISEN AM, MACGUIRE A: Sonographic evaluation of osteoarthritic femoral condylar cartilage. Correlation with operative findings. *Clin Orthop Relat Res* 1990; 254: 230-5.
13. TARHAN S, UNLU Z: Magnetic resonance imaging and ultrasonographic evaluation of the patients with knee osteoarthritis: a comparative study. *Clin Rheumatol* 2003; 22: 181-8.
14. JOSHUA F, LASSERE M, BRUYN GA *et al*.: Summary findings of a systematic review of the ultrasound assessment of synovitis. *J Rheumatol* 2007; 34: 839-47.
15. ABRAHAM AM, GOFF I, PEARCE MS, FRANCIS RM, BIRRELL F: Reliability and validity of ultrasound imaging of features of knee osteoarthritis in the community. *BMC Musculoskelet Disord* 2011; 12: 70.
16. NAREDO E, ACEBES C, MOLLER I *et al*.: Ultrasound validity in the measurement of knee cartilage thickness. *Ann Rheum Dis* 2009; 68: 1322-7.
17. SCHEEL AK, SCHMIDT WA, HERMANN KG *et al*.: Interobserver reliability of rheumatologists performing musculoskeletal ultrasonography: results from a EULAR "Train the trainers" course. *Ann Rheum Dis* 2005; 64: 1043-9.
18. NAREDO E, MOLLER I, MORAGUES C *et al*.: Interobserver reliability in musculoskeletal ultrasonography: results from a "Teach the Teachers" rheumatologist course. *Ann Rheum Dis* 2006; 65: 14-9.
19. ALTMAN R, ASCH E, BLOCH D *et al*.: Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. *Arthritis Rheum* 1986; 29: 1039-49.
20. LANDIS JR, KOCH GG: The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159-74.
21. LIN LI: A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989; 45: 255-68.
22. MANDER A: BATPLOT: Stata module to produce Bland-Altman plots accounting for trend. 2010. Ref Type: Online Source
23. CASTRIOTA-SCANDERBEG A, DE MICHELI V, SCARALE MG, BONETTI MG, CAMMISA M: Precision of sonographic measurement of articular cartilage: inter- and intra-observer analysis. *Skeletal Radiol* 1996; 25: 545-9.
24. SPANNOW AH, PFEIFFER-JENSEN M, ANDERSEN NT, STENBOG E, HERLIN T: Inter- and intraobserver variation of ultrasonographic cartilage thickness assessments in small and large joints in healthy children. *Pediatr Rheumatol Online J* 2009; 7: 12.
25. IAGNOCCO A, MEENAGH G, RIENTE L *et al*.: Ultrasound imaging for the rheumatologist XXIX. Sonographic assessment of the knee in patients with osteoarthritis. *Clin Exp Rheumatol* 2010; 28: 643-6.